Two-Stage Stochastic Semidefinite Programming and Decomposition Based Interior Point Methods: Theory

Sanjay Mehrotra *   M. Gökhan Özevin †

January 5, 2005

Abstract

We introduce two-stage stochastic semidefinite programs with recourse and present a Benders decomposition based linearly convergent interior point algorithms to solve them. This extends the results in Zhao [16] wherein it was shown that the logarithmic barrier associated with the recourse function of two-stage stochastic linear programs with recourse behaves as a strongly self-concordant barrier on the first stage solutions. In this paper we develop the necessary theory. A companion paper [8] addresses implementation issues for the theoretical algorithms of this paper.

Key Words: Stochastic Programming, Semidefinite programming, Benders Decomposition, Interior Point Methods, Primal-Dual Methods

*Corresponding Author, Department of Industrial Engineering & Management Sciences, Northwestern University, Evanston, IL 60208, USA (mehrotra@northwestern.edu).

†Department of Industrial Engineering & Management Sciences, Northwestern University, Evanston, IL 60208, USA (ozevin@northwestern.edu). The research was supported in part by NSF-DMI-0200151 and ONR-N00014-01-1-0048

1. Introduction

We introduce and study the two-stage stochastic semidefinite programming (TSSDP) problem with recourse in the dual standard form:

$$\begin{array}{rll} max & \eta(x) := & c^T x + \rho(\mu, x) \\ & s.t. & Ax + s = b, \\ & & s \in \mathcal{K}^p, \end{array} \tag{1.1}$$

where

$$\rho(x) := E\{\rho(x, \tilde{\xi})\} \tag{1.2}$$

and

$$\begin{array}{rll} \rho(x, \xi) := & max & d(\xi)^T y(\xi) \\ & s.t. & W(\xi) y(\xi) + s(\xi) = h(\xi) - T(\xi) x, \\ & & s(\xi) \in \mathcal{K}^r. \end{array} \tag{1.3}$$

In the first stage problem (1.1), $x \in \mathbb{R}^n$ and $s \in \mathbb{R}^{p^2}$ are decision variables. A is a $p^2 \times n$ matrix with n linearly independent columns that are obtained by vectorization of n symmetric real $p \times p$ matrices and $b \in \mathbb{R}^{p^2}$. We have chosen this form of TSSDP for notational convenience in the analysis of this paper. The cone $\mathcal{K}^\nu := \{\mathbf{vec}(M) \mid M \in \mathbb{R}^{\nu \times \nu}$is symmetric positive semidefinite$\}$, is the cone of vectors obtained from the vectorization of symmetric positive semidefinite matrices. $\mathcal{K}^\nu_+$ is used to describe the cone generated by positive definite matrices.

In (1.2), E represents the expectation with respect to $\tilde{\xi}$ and Ξ is the support of $\tilde{\xi}$. For each realization ξ of $\tilde{\xi}$, $y(\xi) \in \mathbb{R}^m$ and $s(\xi) \in \mathbb{R}^{r^2}$ are decision variables. $h(\xi) \in \mathbb{R}^{r^2}$ and $T(\xi)$ is a $r^2 \times n$ matrix with n linearly independent columns that are obtained by vectorization of n symmetric real $r \times r$ matrices. Similarly, $W(\xi)$ is a $r^2 \times m$ matrix with m linearly independent columns that are obtained by vectorization of m symmetric real $r \times r$ matrices. We assume that Ξ is discrete and finite.

The TSSDP problem is a natural generalization of Semidefinite programming [15] to its two stage stochastic programming counterpart. Problems where objective and constraints are defined by convex quadratic inequalities or second order cone inequalities are special cases. The Linear-quadratic model, which is a special case, was introduced by Rockafellar and Wets [12]. We can write the explicit extensive formulation of this problem, which is a Semi-definite program. We can then solve this extensive formulation directly, in particular, by using primal-dual interior point methods exploiting its special structure through efficient matrix factorization schemes [3, 4, 5]. However, the focus of this paper is in developing decomposition based interior point methods for TSSDP in the spirit of Bender's decomposition. This decomposition approach has several potential advantages because it does not require explicit knowledge of all the scenarios and associated variables in the algorithm. The scenario information is used in gradient and Hessian evaluation central to the algorithm. In practice, this allows for a gradual increase in the number of scenarios during computations as the algorithm progresses. The gradient and Hessian needed to compute the Newton direction is built from the solutions of second stage barrier problems, and their computation decomposes. If information from only a subset of scenarios is used inexact gradients and Hessians is calculated. This may have computational advantages in the single and multi-processor computational environments. In the single processor environment, it may allow for less computations in the early stage of interior point algorithm, particularly when the total number of scenarios is very large. In the multi-processor, and particularly distributed computing environment, where some of the computational nodes may not be reliable, it has the advantage that the algorithm need not depend on completely finishing computations with all the scenarios. Furthermore, decomposition may allow use of information from one scenario to save computational efforts at other scenarios.

In general, the recourse function $\rho(x)$ is not differentiable everywhere. The decomposition approaches either use the nonsmooth optimization techniques [1, 2, 14], or use techniques to smooth this function [12, 13]. Given the success of interior point methods, it is logical to investigate if decomposition based interior point algorithms are possible for stochastic programming problems. Zhao [16] developed a decomposition algorithm by regularizing the second stage problem with a log barrier for linear two stage stochastic programs. In particular, he showed that the log barrier associated with the recourse function of two-stage stochastic linear programs behaves as a strongly self-concordant barrier (see Nesterov and Nemirovskii [9] and Renegar [11]) on the first stage solutions. In this paper we show that the recourse function is also strongly self-concordant for two-stage stochastic semidefinite programs (TSSDP). This allows us to give a Benders decomposition based linearly convergent interior point algorithm for TSSDP. The convergence analysis of this paper forms the conceptual backbone for a more practical algorithm developed an implemented in [8].

This paper is organized as follows. In Section 2 we state our notation, the problem formulation and our assumptions. In Section 3 we show that the barrier recourse function (defined in Section 2) comprises a self-concordant family. In Section 4 we present a conceptual interior point decomposition algorithm and give its convergence theorems. Proofs of these convergence theorems are given in the Section 5.

We use the following notation: For any strictly positive vector x in $\mathbb{R}^n$, we define $x^{-1} := (x_1^{-1}, \ldots, x_n^{-1})^T$. $X := diag(x_1, \ldots, x_n)$ denote the $n \times n$ diagonal matrix whose diagonal entries are $x_1, \ldots, x_n$. An identity matrix of appropriate dimension is denoted by I. Throughout this paper we use "∇","∇^2","∇^3" to denote the gradient, Hessian and the third order derivative with respect to x and a " $'$ " for the derivative with respect to variables other than x. "∇" is also used to denote the Jacobian of a vector function. For example,

$$[\{\nabla^2 f(\mu, x)\}']_{i,j} = \frac{\partial}{\partial \mu}\left(\frac{\partial^2 f(\mu, x)}{\partial x_i \partial x_j}\right).$$

We denote the matrices corresponding to a vector s by $S := \mathbf{mat}(S)$, and $\mathbf{vec(s)}$ is a vector whose elements are the elements of a matrix S. $A \otimes B$ represents the Kronecker product of matrices A and B. The Kronecker product satisfy relationship $[A \otimes B][C \otimes D] = [AC \otimes BD]$, assuming that number of rows in A and B equals the number of columns in C and D. Also, $(A \otimes B)\mathbf{vec}(C) = \mathbf{vec}(BCA^T)$.

2. Problem Formulation and Assumptions

Let the random variable $\tilde{\xi}$ have a finite discrete support $\Xi = \{\xi_1, \ldots, \xi_K\}$ with probabilities $\{\pi_1, \ldots, \pi_K\}$. For simplicity of notation we define $\rho_i(x) := \rho(x, \xi_i)$, $T_i := T(\xi_i)$, $W_i := W(\xi_i)$, $h_i := h(\xi_i)$, $y_i := y(\xi_i)$, and $d_i := \pi_i d(\xi_i)$. The problem (1.1-1.3) is rewritten as

$$\begin{aligned} max \quad \eta(x) := \quad & c^T x + \rho(x) \\ s.t. \quad & Ax + s = b, \\ & s \in \mathcal{K}^p, \end{aligned} \tag{2.1}$$

where

$$\rho(x) := \sum_{i=1}^{K} \rho_i(x) \tag{2.2}$$

and for $i = 1, \ldots, K$

$$\begin{aligned} \rho_i(x) := \quad max \quad & d_i^T y_i \\ s.t. \quad & W_i y_i + s_i = h_i - T_i x, \\ & s_i \in \mathcal{K}^r. \end{aligned} \tag{2.3}$$

Let γ and λ_i be the first and second-stage dual multipliers. The dual of (2.3) is:

$$\begin{aligned} min \quad & (h_i - T_i x)^T \lambda_i \\ s.t. \quad & W_i^T \lambda_i = d_i, \\ & \lambda_i \in \mathcal{K}^r. \end{aligned} \tag{2.4}$$

Here $s_i \in \mathbb{R}^{r^2}$, $W_i \in \mathbb{R}^{r^2 \times m}$, and h_i, T_i is data of appropriate dimensions.

Let us define the following feasibility sets:

$\mathcal{F}_i(x) := \{y_i \mid W_i y_i + s_i = h_i - T_i x,\ s_i \in \mathcal{K}^r\}, \mathcal{F}_i^1 := \{x \mid \mathcal{F}_i(x) \neq \emptyset\}, \mathcal{F}^1 := \cap_{i=1}^K \mathcal{F}_i^1,$
$\mathcal{F}^0 := \mathcal{F}^1 \cap \{x \mid Ax + s = b,\ s \in \mathcal{K}^p\},$ and
$\mathcal{F} := \{(x, s, \gamma) \times (y_1, s_1, \lambda_1, \ldots, y_K, s_K, \lambda_K) \mid Ax + s = b,\ s \in \mathcal{K}^p;\ W_i y_i + s_i = h_i - T_i x,$
$s_i \in \mathcal{K}^r; W_i^T \lambda_i = d_i, \lambda_i \in \mathcal{K}^r\},$ for $i = 1, \ldots, K;\ A^T\gamma + \sum_{i=1}^K T_i^T \lambda_i = c\}.$

We make the following assumption:
A1 $\mathcal{F} \neq \emptyset$, and it has a non-empty relative interior.

A2 Matrices A and W_i have full column rank.

Assumption A1 requires that primal and dual feasible sets of the explicit deterministic equivalent formulation of (2.1–2.3) have non-empty interiors. In particular, it assumes strong duality (see for example, Ramana, Tunçel, Wolkowicz [10]) for first and second stage semidefinite programs. In practice this can be ensured by introducing artificial variables. Assumption A2 is for convenience.

Consider the following log-barrier decomposition problem:

$$\begin{array}{lll} max & \eta(\mu, x) := & c^T x + \rho(x) + \mu \ln \det S \\ & s.t. & Ax + s = b, \\ & & s \in \mathcal{K}^p, \end{array} \tag{2.5}$$

where

$$\rho(\mu, x) := \sum_{i=1}^{K} \rho_i(\mu, x) \tag{2.6}$$

and for $i = 1, \ldots, K$

$$\begin{array}{lll} \rho_i(\mu, x) := & max & d_i^T y_i + \mu \ln \det S_i \\ & s.t. & W_i y_i + s_i = h_i - T_i x, \\ & & s_i \in \mathcal{K}^r. \end{array} \tag{2.7}$$

The log-barrier problem associated with the dual (2.4) is given by:

$$\begin{array}{ll} min & (h_i - T_i x)^T \lambda_i - \mu \ln \det \Lambda_i \\ s.t. & W_i^T \lambda_i = d_i, \\ & \lambda_i \in \mathcal{K}^r. \end{array} \tag{2.8}$$

Note that for a given $\mu > 0$, the log-barrier recourse function $\rho(\mu, x) < \infty$ iff $x \in \mathcal{F}^1$. Hence it describes the interior of $\mathcal{F}^0$ implicitly. Assumption A1 implies that each of the problems (2.5, 2.7–2.10) below have a unique solution. Since problems (2.7) and (2.8) are respectively concave and convex, (y_i, s_i) and λ_i are optimal solutions to (2.7) and (2.8), respectively, if and only if they satisfy the following optimality conditions:

$$\begin{array}{l} W_i^T \lambda_i = d_i, \\ W_i y_i + s_i = h_i - T_i x, \\ S_i \Lambda_i = \mu I, \\ \lambda_i, s_i \in \mathcal{K}^r_+. \end{array} \tag{2.9}$$

Note that $\Lambda_i = \mathbf{vec}(\lambda_i)$. Throughout the paper we denote the optimal solution of the first stage problem (2.5) by $x(\mu)$ and the solutions of the optimality conditions (2.9) for a given $x \in \mathcal{F}^1$ by $(y_i(\mu,x), s_i(\mu,x), \lambda_i(\mu,x))$.
The optimal solutions of (2.5-2.7) and those of the extensive log-barrier problem:

$$
\begin{array}{ll}
min & c^T x + \sum_{i=1}^{K} d_i^T y_i + \mu \ln\det S + \mu e^T \ln\det S_i \\
s.t. & Ax + s = b, \\
 & W_i y_i + s_i = h_i - T_i x, \quad i = 1, \ldots, K, \\
 & s \in \mathcal{K}^p, \;\; s_i \in \mathcal{K}^r, \quad i = 1, \ldots, K,
\end{array}
\tag{2.10}
$$

associated with the extensive formulation of (2.1-2.3) have the following relationship.

Proposition 2.1 *For a given $\mu > 0$, if $(x(\mu), s(\mu); y_1(\mu), s_1(\mu), \ldots, y_K(\mu), s_K(\mu))$ is the optimal solution of (2.10), then $(x(\mu), s(\mu))$ is the optimal solution of (2.5), and $(y_1(\mu), s_1(\mu), \ldots, y_K(\mu), s_K(\mu))$ are the optimal solutions of subproblems (2.7) for the given μ and $x = x(\mu)$. Conversely, if for given μ, $(x(\mu), s(\mu))$ is the optimal solution of (2.5) and $(y_1(\mu), s_1(\mu), \ldots, y_K(\mu), s_K(\mu))$ are the optimal solutions of (2.7) with $x = x(\mu)$, then $(x(\mu), s(\mu)$ $y_1(\mu), s_1(\mu), \ldots, y_K(\mu), s_K(\mu))$ is the optimal solution of (2.10).* $\square$

3. The Self-Concordance Properties of the Log-Barrier Recourse.

3.1 Computation of $\nabla\eta(\mu, x)$ and $\nabla^2\eta(\mu, x)$

From (2.9) we can show that the optimal objective values of primal and dual barrier problems (2.7–2.8) differ by a constant term, in particular:

$$
\rho_i(\mu, x) = (h_i - T_i x)\lambda_i(\mu, x) - \mu \ln\det \Lambda_i(\mu, x) + r\mu(1 - \ln \mu). \tag{3.1}
$$

In order to compute $\nabla\eta(\mu, x)$ and $\nabla^2\eta(\mu, x)$ in (3.8) we need to determine the derivative of $\lambda_i(\mu, x)$ with respect to x. Let $(y_i, \lambda_i, s_i) := (y_i(\mu, x), \lambda_i(\mu, x), s_i(\mu, x))$. Differentiating (2.9) with respect to x we obtain

$$
\begin{array}{rcl}
W_i^T \nabla\lambda_i & = & 0, \\
W_i \nabla y_i + \nabla s_i & = & -T_i, \\
(I \otimes S_i)\nabla\lambda_i + (\Lambda_i \otimes I)\nabla s_i & = & 0.
\end{array}
\tag{3.2}
$$

Solving the system (3.2) we get

$$
\begin{array}{rcl}
\nabla y_i & = & -R_i^{-1} W_i^T Q_i^2 T_i, \\
\nabla\lambda_i & = & Q_i P_i Q_i T_i, \\
\nabla s_i & = & -Q_i^{-1} P_i Q_i T_i,
\end{array}
\tag{3.3}
$$

where

$$Q_i := Q_i(\mu, x) = (\Lambda_i \otimes S_i^{-1})^{1/2}, \qquad R_i := R_i(\mu, x) = W_i^T Q_i^2 W_i \tag{3.4}$$

$$\text{and } P_i := P_i(\mu, x) = I - Q_i W_i R_i^{-1} W_i^T Q_i. \tag{3.5}$$

Now differentiating (3.1) and using the optimality conditions (2.9) and (3.3) we can verify that

$$\nabla \rho_i(\mu, x) = -T_i^T \lambda_i(\mu, x), \text{ and } \nabla^2 \rho_i(\mu, x) = -T_i^T \nabla \lambda_i(\mu, x). \tag{3.6}$$

Hence,

$$\nabla \eta(\mu, x) = c - \sum_{i=1}^{K} T_i^T \lambda_i(\mu, x) - \mu A^T s^{-1}, \tag{3.7}$$

$$\nabla^2 \eta(\mu, x) = -\sum_{i=1}^{K} T_i^T \nabla \lambda_i(\mu, x) - \mu A^T (S^{-1} \otimes S^{-1}) A. \tag{3.8}$$

Then, substituting for $\nabla \lambda_i$ in (3.8) we get

$$\nabla^2 \eta(\mu, x) = -\sum_{i=1}^{K} T_i^T Q_i P_i Q_i T_i - \mu A^T (S^{-1} \otimes S^{-1}) A. \tag{3.9}$$

3.2 Self-Concordance of the Recourse Function

The following definition of self-concordant functions is introduced by Nesterov and Nemirovskii [9].

Definition 3.1 *(Nesterov and Nemirovskii [9]) Let $\mathcal{E}$ be a finite-dimensional real vector space, $\mathcal{Q}$ be an open nonempty convex subset of $\mathcal{E}$, $f : \mathcal{Q} \to \mathbb{R}$ be a function , $\alpha > 0$. f is called α-self-concordant on $\mathcal{Q}$ with the parameter value α, if $f \in C^3$ is a convex function on $\mathcal{Q}$, and, for all $x \in \mathcal{Q}$ and $h \in \mathcal{E}$ the following inequality holds:*

$$|\nabla^3 f(x)[h, h, h]| \leq 2\alpha^{-1/2} (\nabla^2 f(x)[h, h])^{3/2}.$$

An α-self-concordant on $\mathcal{Q}$ function f is called strongly α-self-concordant on $\mathcal{Q}$ if $f(x_i)$ tends to infinity along every sequence $\{x_i \in \mathcal{Q}\}$ converging to a boundary point of $\mathcal{Q}$.

We now show that recourse function $\rho(\mu, x)$ behaves as a strongly self-concordant barrier on $\mathcal{F}^1$.

Lemma 3.1 *For any fixed $\mu > 0$, $\rho_i(\mu, \cdot)$ is strongly μ-self-condordant on $\mathcal{F}_i^1, i = 1, \ldots, K$.*

Proof. For any $\mu > 0$, $d \in \mathbb{R}^n$ and $\bar{x} \in \{x \mid \rho_i(x) < \infty\}$ we define the univariate function

$$\Phi_i(t) := \nabla^2 \rho_i(\mu, \bar{x} + td)[d, d].$$

Note that $\Phi_i'(0) = \nabla^3 \rho_i(\mu, \bar{x})[d, d, d]$. Along every sequence $\{x_j \in \mathcal{F}_i^1\}$ converging to the boundary of $\mathcal{F}_i^1$, $\rho_i(\mu, x_j)$ tends to infinity. To prove this lemma it suffices to show that

$$|\Phi_i(0)'| \leq \frac{2}{\sqrt{\mu}} |\Phi_i(0)|^{3/2}.$$

Let $(\lambda_i(t), P_i(t), s_i(t), Q_i(t), R_i(t)) := (\lambda_i(\mu, \bar{x} + td), P_i(\mu, \bar{x} + td), s_i(\mu, \bar{x} + td), Q_i(\mu, \bar{x} + td), R_i(\mu, \bar{x} + td))$. We define $u_i(t) := P_i(t)Q_i(t)T_i(t)d$. The argument '$(t)$' is dropped when considering all of these variables and their derivatives at $t = 0$, e.g., $u' := u'(0)$. Note that $\Phi_i(0) = -u_i^T u_i = -\|u_i\|^2$ and thus $|\Phi_i(0)'| = |2u_i^T u_i'|$.

The first equality below following from using (3.4,3.5). The second equality is derived by using derivatives by parts. The third equality uses $R_i' = W_i^T[Q_i Q_i' + Q_i' Q_i]W_i$.
We have,

$$\begin{aligned}
u_i' &= [Q_i - Q_i W_i R_i^{-1} W_i Q_i^2]' T_i d \\
&= [Q_i' - Q_i' W_i R_i^{-1} W_i Q_i^2 + Q_i W_i R_i^{-1} R_i' R_i^{-1} W_i Q_i^2 - Q_i W_i R_i^{-1} W_i (Q_i Q_i' + Q_i' Q_i)] T_i d \\
&= [Q_i'(I - W_i R_i^{-1} W_i Q_i^2) - Q_i W_i R_i^{-1} W_i (Q_i Q_i' + Q_i' Q_i)(I - W_i R_i^{-1} W_i Q_i^2)] T_i d \\
&= [(Q_i' - Q_i W_i R_i^{-1} W_i (Q_i Q_i' + Q_i' Q_i))](I - W_i R_i^{-1} W_i Q_i^2) T_i d \\
&= [(Q_i' - Q_i W_i R_i^{-1} W_i (Q_i Q_i' + Q_i' Q_i))] Q_i^{-1} u_i \quad \text{(noting that } (I - W_i R_i^{-1} W_i Q_i^2) T_i d = Q_i^{-1} u_i)
\end{aligned} \tag{3.10}$$

Observing that $u_i^T Q_i W_i = 0$, from (3.10) we get

$$\begin{aligned}
|\Phi_i(0)'| &= |2u_i^T u_i'| = |2u_i^T Q_i' Q_i^{-1} u_i| \\
&= |u_i^T (Q_i' Q_i^{-1} + Q_i^{-1} Q_i') u_i| \quad \text{(since } Q_i, Q_i' \text{ are symmetric matrices)} \\
&= |u_i^T Q_i^{-1} (Q_i Q_i' + Q_i' Q_i) Q_i^{-1} u_i| = |u_i^T Q_i^{-1} (Q_i^2)' Q_i^{-1} u_i|.
\end{aligned} \tag{3.11}$$

We let $\nabla \lambda_i := \nabla \lambda_i(\mu, \bar{x})$ and $\lambda_i' := \left.\frac{\partial \lambda_i(\mu, \bar{x} + td)}{\partial t}\right|_{t=0} = \nabla \lambda_i d$. Note that from (3.4) we have

$$\begin{aligned}
(Q_i^2)' &= (\Lambda_i \otimes S_i^{-1})' = \mu^{-1}(\Lambda_i \otimes \Lambda_i)' = \mu^{-1}(\Lambda_i \otimes \Lambda_i' + \Lambda_i' \otimes \Lambda_i) \quad \text{(since } \Lambda_i' = \mathbf{mat}(\nabla \lambda_i d)) \\
&= \mu^{-1}(\Lambda_i \otimes \mathbf{mat}(\nabla \lambda_i d) + \mathbf{mat}(\nabla \lambda_i d) \otimes \Lambda_i).
\end{aligned} \tag{3.12}$$

Combining (3.11), (3.12) and using (3.4) we obtain,

$$\begin{aligned}
|\Phi_i(0)'| &= |u_i^T(\Lambda_i^{-1/2}\otimes\Lambda_i^{-1/2})[(\Lambda_i\otimes\mathbf{mat}(\nabla\lambda_i d)+\mathbf{mat}(\nabla\lambda_i d)\otimes\Lambda_i)](\Lambda_i^{-1/2}\otimes\Lambda_i^{-1/2})u_i| \\
&= |u_i^T[I\otimes(\Lambda_i^{-1/2}\mathbf{mat}(\nabla\lambda_i d)\Lambda_i^{-1/2})+(\Lambda_i^{-1/2}\mathbf{mat}(\nabla\lambda_i d)\Lambda_i^{-1/2})\otimes I]u_i| \\
&\le 2\|u_i\|_2^2\ \|\mathbf{vec}(\Lambda_i^{-1/2}\mathbf{mat}(\nabla\lambda_i d)\Lambda_i^{-1/2})\|_2 \\
&= 2\|u_i\|_2^2\ \|(\Lambda_i^{-1/2}\otimes\Lambda_i^{-1/2})(\nabla\lambda_i d)\|_2 \\
&= 2\mu^{-1/2}\|u_i\|_2^2\ \|Q_i^{-1}\nabla\lambda_i d\|_2 \qquad \text{(noting that } Q_i^{-1}=\sqrt{\mu}(\Lambda_i^{-1/2}\otimes\Lambda_i^{-1/2})) \\
&= 2\mu^{-1/2}\|u_i\|_2^3 \qquad \text{(noting that } Q_i^{-1}\nabla\lambda_i d=u_i) \\
&= 2\mu^{-1/2}|\Phi_i(0)|^{3/2} \qquad \text{(since } |\Phi_i(0)|=\|u_i\|_2^2).\ \square
\end{aligned} \tag{3.13}$$

We have the following corollary.

Corollary 3.1 *The recourse function $\rho(\mu,x)$ is a μ-self-concordant barrier on $\mathcal{F}^1$ and the first stage objective function $\eta(\mu,x):=c^Tx+\rho(x)+\mu\ln\det S$ is a strongly μ-self-concordant barrier on $\mathcal{F}^0$.*

Proof. It is easy to verify that $\mu\ln\det S$ is strongly μ-self-concordant barrier on $\{x|Ax+s=b, s\in\mathcal{K}^p\}$. The corollary follows from Proposition 2.1.1 (ii) in [9]. $\square$

3.3 Parameters of the Self-Concordant Family

The self-concordant family with appropriate parameters is defined in Nesterov and Nemirovskii [9]. They showed that given such a family, the parameters defining the family allow us to relate the rate at which the barrier parameter μ is varied and the number of Newton steps required to maintain the proximity to the central path. Below is the definition of a strongly self-concordant family adapted to the current setting from the original definition in Nesterov and Nemirovskii [9]. These conditions might look rather technical; nevertheless they simplify our convergence analysis and the accompanying proofs in the sequel and explicitly reveal some essential properties of the log-barrier recourse function $\rho(\mu,x)$. They allow us to invoke the interior point convergence theory developed by Nesterov and Nemirovskii [9].

Definition 3.2 *The family of functions $\{\eta(\mu,\cdot):\mu>0\}$ is strongly self-concordant on $\mathcal{F}^0$ with parameter functions $\alpha(\mu),\gamma(\mu),\nu(\mu),\xi(\mu)$, and $\sigma(\mu)$ if*

1. *If $\eta(\mu,x)$ is concave in x, continuous in $(\mu,x)\in\mathbb{R}_{++}\times\mathcal{F}^0$ and has three derivatives in x, continuous in $(\mu,x)\in\mathbb{R}_{++}\times\mathcal{F}^0$.*
2. *$\nabla\eta(\mu,x)$ and $\nabla^2\eta(\mu,x)$ are continuously differentiable in μ,*
3. *For any $\mu\in\mathbb{R}_{++}$, $\eta(\mu,x)$ is strongly $\alpha(\mu)$-self-concordant on $\mathcal{F}^0$,*

4. *The parameter functions $\alpha(\mu), \gamma(\mu), \xi(\mu)$ and $\sigma(\mu)$ are continuous positive scalar functions on $\mu \in \mathbb{R}_{++}$,*

5. *For every $(\mu, x) \in \mathbb{R}_{++} \times \mathcal{F}^0$ and $h \in \mathbb{R}^n$,*

$$|\{\nabla\eta(\mu, x)h\}' - \{\ln \nu(\mu)\}'\{\nabla\eta(\mu, x)h\}| \leq \xi(\mu)\alpha(\mu)^{1/2}(-h^T\nabla^2\eta(\mu, x)h)^{1/2},$$

6. *For every $(\mu, x) \in \mathbb{R}_{++} \times \mathcal{F}^0$ and $h \in \mathbb{R}^n$,*

$$|\{h^T\nabla^2\eta(\mu, x)h\}' - \{\ln \gamma(\mu)\}'h^T\nabla^2\eta(\mu, x)h| \leq -2\sigma(\mu)h^T\nabla^2\eta(\mu, x)h.$$

We refer the reader to Nesterov and Nemirovskii [9] for the original definition of self-concordant families and their properties. The essence of the above definition is in conditions 5 and 6.

Theorem 3.1 *The family of functions $\eta : \mathbb{R}_{++} \times \mathcal{F} \mapsto \mathbb{R}$ is a strongly self-concordant family with parameters $\alpha(\mu) = \mu$, $\gamma(\mu) = \nu(\mu) = 1$, $\xi(\mu) = \frac{\sqrt{p+Kr}}{\mu}$ and $\sigma(\mu) = \frac{\sqrt{r}}{2\mu}$.*

Proof. It is easy to verify that conditions 1 through 4 of Definition 3.2 hold. Lemma 3.2 and Lemma 3.3 below show that conditions 5 and 6 are satisfied. □

In Lemmas 3.2 and 3.3 we bound the changes of $\nabla\eta(\mu, x)$ and $\nabla^2\eta(\mu, x)$ as the barrier parameter μ changes. This requires us to calculate (y_i', λ_i', s_i'), which are the derivatives of $(y_i(\mu, x), \lambda_i(\mu, x), s_i(\mu, x))$ with respect to μ. Differentiating (2.9) with respect to μ we get

$$\begin{aligned} W_i^T\lambda_i' &= 0, \\ W_iy_i' + s_i' &= 0, \\ (I \otimes S_i)\lambda_i' + (\Lambda_i \otimes I)s_i' &= \mathbf{vec}(I). \end{aligned} \tag{3.14}$$

Solving (3.14) we obtain

$$\begin{aligned} y_i' &= -R_i^{-1}W_i^Ts_i^{-1}, \\ \lambda_i' &= \frac{1}{\sqrt{\mu}}Q_iP_i\mathbf{vec}(I), \\ s_i' &= W_iR_i^{-1}W_i^Ts_i^{-1}. \end{aligned} \tag{3.15}$$

Lemma 3.2 *For any $\mu > 0$, $x \in \mathcal{F}^0$ and $h \in \mathbb{R}^n$ we have*

$$|\{\nabla\eta(\mu, x)^Th\}'| \leq \left[\frac{-(p+Kr)}{\mu}h^T\nabla^2\eta(\mu, x)^Th\right]^{1/2}.$$

Proof. Differentiating (3.7) with respect to μ and applying (3.15) we get

$$\begin{aligned}\{\nabla\eta(\mu,x)\}' &= -\frac{1}{\sqrt{\mu}}\sum_{i=1}^{K} T_i^T Q_i P_i \mathbf{vec}(I) - A^T s^{-1} \\ &= -\frac{1}{\sqrt{\mu}}\sum_{i=1}^{K} T_i^T Q_i P_i \mathbf{vec}(I) - A^T (S^{-1/2}\otimes S^{-1/2})\mathbf{vec}(I).\end{aligned}$$

We define

$$B := \left[\frac{1}{\sqrt{\mu}} T_1^T Q_1 P_1,\ \ldots,\ \frac{1}{\sqrt{\mu}} T_K^T Q_K P_K,\ A^T(S^{-1/2}\otimes S^{-1/2})\right],$$

and let z be an $(p^2 + Kr^2)$ dimensional vector defined by $z := [\mathbf{vec}(I_r), \ldots,\ \mathbf{vec}(I_r), \mathbf{vec}(I_p)]$. We can write

$$\{\nabla\eta(\mu,x)\}' = -Bz. \tag{3.16}$$

Note that $BB^T = \frac{1}{\mu}\sum_{i=1}^{K} T_i^T Q_i P_i Q_i T_i + A^T(S^{-1}\otimes S^{-1})A = -\frac{1}{\mu}\nabla^2\eta(\mu,x)$.

Now we have

$$-\{\nabla\eta(\mu,x)^T\}'[\nabla^2\eta(\mu,x)]^{-1}\{\nabla\eta(\mu,x)\}' = \frac{1}{\mu} z^T B^T [BB^T]^{-1} Bz \le \frac{1}{\mu} z^T z = \frac{1}{\mu}(p+Kr). \tag{3.17}$$

Now by using norm inequalities and (3.17) it follows that

$$\begin{aligned}|\{\nabla\eta(\mu,x)^T h\}'| &\le \left[-\{\nabla\eta(\mu,x)^T\}'[\nabla^2\eta(\mu,x)]^{-1}\{\nabla\eta(\mu,x)\}'\right]^{1/2}\left[-h^T\nabla^2\eta(\mu,x)h\right]^{1/2} \\ &\le \left[\frac{-(p+Kr)}{\mu} h^T\nabla^2\eta(\mu,x)h\right]^{1/2}.\ \square\end{aligned}$$

Lemma 3.3 *For any $\mu > 0$, $x \in \mathcal{F}^0$ and $h \in \mathbb{R}^n$ we have*

$$|\{h^T\nabla^2\eta(\mu,x)h\}'| \le -\frac{\sqrt{r}}{\mu} h^T\nabla^2\eta(\mu,x)h.$$

Proof. We fix $h \in \mathbb{R}^n$ and let $(\lambda_i, P_i, s_i, Q_i, R_i) := (\lambda_i(\mu,x), P_i(\mu,x), s_i(\mu,x), Q_i(\mu,x), R_i(\mu,x))$. Let us denote $u_i := P_i Q_i T_i d$. We have

$$h^T\nabla^2\eta(\mu,x)h = -\sum_{i=1}^{K} u_i^T u_i - \mu h^T A^T (S^{-1}\otimes S^{-1})Ah.$$

From the proof of Lemma 3.1 (see 3.11), we have

$$\{h^T \nabla^2 \eta(\mu, x) h\}' = -\sum_{i=1}^{K} u_i^T Q_i^{-1} (Q_i^2)' Q_i^{-1} u_i - h^T A^T (S^{-1} \otimes S^{-1}) A h. \tag{3.18}$$

From (3.15), definition of Q_i from (3.4), $S_i \Lambda_i = \mu I$, and using $S_i \Lambda_i' + \Lambda_i S_i' = I$, $\Lambda_i^{-1} = \mu^{-1} S_i$, it follows that

$$\begin{aligned}
u_i^T Q_i^{-1} (Q_i^2)' Q_i^{-1} u_i &= u_i^T (I \otimes \Lambda_i^{-1/2} \Lambda_i' \Lambda_i^{-1/2} - S_i^{-1/2} S_i' S_i^{-1/2} \otimes I) u_i \\
&= \mu^{-1} u_i^T (I \otimes I - \mu (S_i^{-1/2} S_i' S_i^{-1/2} \otimes I + I \otimes S_i^{-1/2} S_i' S_i^{-1/2}) u_i \\
&\leq \frac{\|u_i\|_2^2}{\mu} \|\mathbf{vec}(I) - 2\mu (S_i^{-1/2} \otimes S_i^{-1/2}) s_i'\|_2 \\
&= \frac{\|u_i\|_2^2}{\mu} \|\mathbf{vec}(I) - 2\mu (S_i^{-1/2} \otimes S_i^{-1/2}) W_i R_i^{-1} W_i^T s_i^{-1}\|_2 \\
&= \frac{\|u_i\|_2^2}{\mu} \|(I - 2P_i) \mathbf{vec}(I)\|_2 \\
&\leq \frac{\sqrt{r}}{\mu} \|u_i\|_2^2 \qquad (\text{since } I - 2P \preceq I, \|(I - 2P_i)\|_2 \leq 1).
\end{aligned} \tag{3.19}$$

From (3.18) and (3.19), we obtain for any $h \in R^n$

$$|\{h^T \nabla^2 \eta(\mu, x) h\}'| \leq \frac{\sqrt{r}}{\mu} \sum_{i=1}^{K} u_i^T u_i + h^T A^T (S^{-1} \otimes S^{-1}) A h = -\frac{\sqrt{r}}{\mu} h^T \nabla^2 \eta(\mu, x) h. \quad \square$$

4. The Two-Stage Stochastic SDP Algorithm

Once it is established that the family of functions $\{\eta(\mu, \cdot) : \mu > 0\}$ is strongly self concordant the development of primal path following interior point methods is straight forward. These methods reduce μ by a factor at each iteration and seek to approximate the minimizer $x(\mu)$ for each μ by taking one or more Newton steps. The novelty of the algorithm in the context of TSSDP is in computing the Newton direction from the solutions of the decomposed second stage problems. As μ varies, the minimizers $x(\mu)$ form the central path. By tracing the central path as $\mu \to 0$ this procedure will generate a strictly feasible ϵ-solution to (2.5).

For a given μ the optimality condition for the problem (2.5) is:

$$\nabla \eta(\mu, x(\mu)) = 0. \tag{4.1}$$

Hence at a feasible point x the Newton direction is given by

$$\Delta x = -\nabla \eta(\mu, x)^{-2} \nabla \eta(\mu, x). \tag{4.2}$$

Note that although problems (2.5- 2.7) and (2.10) share the same central path, the associated Newton directions are not identical and lead to different ways of path following. A conceptual primal path following algorithm is given below.

4.1 Conceptual Algorithm

Here $\beta > 0, \gamma \in (0,1)$ and $\theta > 0$ are suitable scalars. We make their values more precise in Theorems 4.1 and 4.2. The desired precision ϵ, an initial point $x^0 \in \mathcal{F}^0$ and μ^0 are given as inputs.

Initialization.
$x = x^0$; $\mu = \mu^0$.

Step 1.

1.1. For all i solve the optimality conditions (2.9) to find $(y_i(\mu, x), s_i, \lambda_i(\mu, x)$.

1.2. Compute the Newton direction Δx from (4.2).

1.3. Let $\delta(\mu, x) = \sqrt{-\frac{1}{\mu}\Delta x^T \nabla^2 \eta(\mu, x)\Delta x}$. If $\delta \leq \beta$ go to Step2.

1.4. Set $x = x + \theta \Delta x$ and go to Step 1.1.

Step 2. If $\mu \leq \epsilon$ stop, otherwise set $\mu = \gamma\mu$ and go to Step 1.1.

In the above algorithm we assume that we can find exact solutions of the optimality conditions (2.9). This assumption considerably simplifies the complexity analysis. In the practical implementation of this algorithm we use approximate solutions of the optimality conditions (2.9) to construct the Newton direction (4.2).

Theorems 4.1 and 4.2 give two standard complexity results for the generic primal interior point method. In the short-step version of the algorithm barrier parameter μ is decreased by a factor $1 - \sigma/\sqrt{n+m}$ $(\sigma > 0)$ in each iteration.

An iteration of the short-step algorithm is performed as follows. At the beginning of iteration k, x^k is close to the central path, i.e. $\delta(\mu^k, x^k) \leq \beta$. After reducing the parameter from μ^k to $\mu^{k+1} = \gamma\mu^k$, we will have $\delta(\mu^{k+1}, x^k) \leq 2\beta$. Then a Newton step with step size $\theta = 1$ is taken resulting in a new point x^{k+1} with $\delta(\mu^{k+1}, x^{k+1}) \leq \beta$. We have the following theorem.

Theorem 4.1 *Let μ^0 be the initial barrier parameter, $\epsilon > 0$ the stopping criterion and $\beta = (2-\sqrt{3})/2$. If the starting point x^0 is sufficiently close to the central path, i.e. $\delta(\mu^0, x^0) \leq \beta$, then the short-step algorithm reduces the barrier parameter μ at a linear rate and terminates within $O(\sqrt{p+Kr}\ln \mu^0/\epsilon)$ iterations.*

Proof: See Section 5.1.

In the long-step version we decrease the barrier parameter μ by an arbitrarily constant factor ($\lambda \in (0,1)$). It has potential for much faster progress, however, several damped Newton steps might be needed for restoring the proximity to the central path. We have the following theorem.

Theorem 4.2 *Let μ^0 be the initial barrier parameter and $\epsilon > 0$ be the stopping criterion and $\beta = 1/6$. If the starting point x^0 is sufficiently close to the central path, i.e. $\delta(\mu^0, x^0) \leq \beta$, then the long-step algorithm reduces the barrier parameter μ at a linear rate and terminates within $O((p + Kr)\ln \mu^0/\epsilon)$ iterations.*

Proof: See Section 5.2.

5. Convergence Proof for Short and Long Step Algorithms

Part (i) of the following proposition follows directly from the definition of self-concordance and is due to Nesterov and Nemirovskii [9, Theorem 2.1.1]. Part (ii) is a corollary of part (i) and is given in Zhao [16] without a proof.

Proposition 5.1 *For any $\mu > 0$, $x \in \mathcal{F}^0$ and Δx we let $\delta := \sqrt{-\frac{1}{\mu}\Delta x^T \nabla^2 \eta(\mu, x)\Delta x}$. Then for $\delta < 1$, $\tau \in [0,1]$ and any $h \in \mathbb{R}^n$ we have*

$$(i) \quad -(1-\tau\delta)^2 h^T \nabla^2 \eta(\mu, x) h \leq -h^T \nabla^2 \eta(\mu, x + \tau\Delta x) h \leq -(1-\tau\delta)^{-2} h^T \nabla^2 \eta(\mu, x) h,$$

$$(ii) \quad |h_1^T [\nabla^2 \eta(\mu, x + \tau\Delta x) - \nabla^2 \eta(\mu, x)] h_2| \leq \\ [(1-\tau\delta)^{-2} - 1]\sqrt{-h_1^T \nabla^2 \eta(\mu, x) h_1}\sqrt{-h_2^T \nabla^2 \eta(\mu, x) h_2}.$$

For the estimation of number of Newton steps needed for recentering we use two different merit functions to measure the speed of Newton's method. We use $\delta(\mu, x)$ for the short-step algorithm and the first stage objective $\eta(\mu, x)$ (defined in Step 1) for the long-step algorithm. The following lemma is due to Theorem 2.2.3 in [9] and describes the behavior of the Newton method as applied to $\eta(\mu, \cdot)$.

Lemma 5.1 *Let $\mu > 0$ and $x \in \mathcal{F}^0$. Furthermore, let Δx be the Newton direction calculated by (4.2) and $\delta(\mu, x) := \sqrt{-\frac{1}{\mu}\Delta x^T \nabla^2 \eta(\mu, x)\Delta x}$. Then the following relations hold:*

(i) If $\delta < 2 - \sqrt{3}$ then

$$\delta(\mu, x + \Delta x) \leq \left(\frac{\delta}{1-\delta}\right)^2 \leq \frac{\delta}{2}.$$

(ii) If $\delta \geq 2 - \sqrt{3}$ *then*

$$\eta(\mu, x) - \eta(\mu, x + \overline{\theta}\Delta x) \geq \mu[\delta - \ln(1 + \delta)],$$

where $\overline{\theta} = (1 + \delta)^{-1}$.

5.1 Complexity of the Short-Step Algorithm

We now show that in this version of the algorithm a single Newton step is sufficient for recentering after updating the barrier parameter μ. To this end we make use of Theorem 3.1.1 in [9], which is restated for the present context in the next proposition.

Proposition 5.2 *Let* $\varphi_\kappa(\eta; \mu, \mu^+) := \left(\frac{1+r}{2} + \frac{\sqrt{p+Kr}}{\kappa}\right) \ln \gamma^{-1}$. *Assume that* $\delta(\mu, x) < \kappa$ *and* $\mu^+ := \gamma\mu$ *satisfies*

$$\varphi_\kappa(\eta; \mu, \mu^+) \leq 1 - \frac{\delta(\mu, x)}{\kappa}.$$

Then $\delta(\mu^+, x) < \kappa$.

Lemma 5.2 *Let* $\mu^+ = \gamma\mu$ *where* $\gamma = 1 - \sigma/\sqrt{p + Kr}$ *and* $\sigma \leq 0.1$. *Furthermore let* $\beta = (2 - \sqrt{3})/2$. *If* $\delta(\mu, x) \leq \beta$ *then* $\delta(\mu^+, x) \leq 2\beta$.

Proof. Let $\kappa = 2\beta = 2 - \sqrt{3}$. It is easy to verify that with $\sigma \leq 0.1$, μ^+ satisfies

$$\begin{aligned} \varphi_\kappa(\eta; \mu, \mu^+) &= \left(\frac{1+r}{2} + \frac{\sqrt{p+Kr}}{\kappa}\right) \ln(1 - \sigma/\sqrt{p + Kr})^{-1} \\ &\leq \frac{1}{2} \leq 1 - \frac{\delta(\mu, x)}{\kappa}. \end{aligned}$$

Now Proposition 5.2 implies

$$\delta(\mu^+, x) \leq \kappa = 2\beta. \quad \square$$

From Lemma 5.1 and Lemma 5.2 it is clear that we can reduce μ by the factor $\gamma = 1 - \sigma/\sqrt{p + Kr}$, $\sigma < 0.1$ at each iteration and a single Newton step is sufficient to restore proximity to the central path.

Hence Theorem 4.1 follows.

5.2 Complexity of the Long-Step Algorithm

For the analysis of the long-step algorithm we use η as the merit function since the iterates generated by the less conservative long-step algorithm may violate the condition, $\delta < 2 - \sqrt{3}$, required in part (i) of Lemma 5.1. Our analysis follows the steps in Zhao [16].

Assume that we have a point x^{k-1} sufficiently close to $x(\mu^{k-1})$. Then we reduce the barrier parameter from μ^{k-1} to $\mu^k = \gamma\mu^{k-1}$, where $\gamma \in (0,1)$. While searching for a point x^k that is sufficiently close to $x(\mu^k)$ the long-step algorithm generates a finite sequence of points (inner iterates) $p_1, \ldots, p_N \in \mathcal{F}^0$, and we finally set $x^k = p_N$. We need to determine an upper bound on N, the number of Newton iteration needed for recentering. We begin by determining an upper bound on the difference

$$\phi(\mu^k, x^{k-1}) := \eta(\mu^k, x(\mu^k)) - \eta(\mu^k, x^{k-1}). \tag{5.1}$$

Then by part (ii) of Lemma 5.1 we know that at $p_i \in \mathcal{F}^0$, independent of i, a Newton step with step size $\overline{\theta} = (1+\delta)^{-1}$ decreases $\eta(\mu^k, p_i)$ at least by a certain amount which depends on the current value of δ and μ. A line search might yield an even larger decrease, however, performing such a line search may be expensive. The theoretical analysis gives an upper bound on N.

The next lemma gives upper bounds on $\phi(\mu^{k-1}, x)$ and $\phi'(\mu^{k-1}, x)$, respectively, for any $\mu > 0$ and $x \in \mathcal{F}^0$. They facilitate us bounding $\phi(\mu^k, x)$.

Lemma 5.3 *Let $\mu > 0$ and $x \in \mathcal{F}^0$. We denote $\tilde{\Delta}x := x - x(\mu)$ and define*

$$\tilde{\delta}(\mu, x) := \sqrt{-\frac{1}{\mu}\tilde{\Delta}x^T \nabla^2 \eta(\mu, x)\tilde{\Delta}x}.$$

For any $\mu > 0$ and $x \in \mathcal{F}^0$, if $\tilde{\delta} < 1$, then

$$\phi(\mu, x) \le \mu\left[\frac{\tilde{\delta}}{1-\tilde{\delta}} + \ln(1-\tilde{\delta})\right], \tag{5.2}$$

$$|\phi'(\mu, x)| \le -\sqrt{p+Kr}\,\ln(1-\tilde{\delta}). \tag{5.3}$$

Proof.

$$\phi(\mu, x) = \eta(\mu, x(\mu)) - \eta(\mu, x) = \int_1^0 \nabla\eta(\mu, x - (1-\tau)\tilde{\Delta}x d\tau.$$

Since $x(\mu)$ is the optimal solution of (2.5), it satisfies the optimality conditions (4.1). Using (4.1) we have

$$\begin{aligned}
\phi(\mu, x) &= \int_1^0 \int_0^\tau \tilde{\Delta}x^T \nabla^2 \eta(\mu, x - (1-\alpha)\tilde{\Delta}x)\tilde{\Delta}x d\alpha d\tau \\
&\le \int_0^1 \int_0^\tau \frac{\mu\tilde{\delta}^2}{(1-(1-\alpha)\tilde{\delta})^2} d\alpha d\tau \qquad \text{(using Proposition 5.1 (i))} \\
&= \mu\left[\frac{\tilde{\delta}}{1-\tilde{\delta}} + \ln(1-\tilde{\delta})\right]. \qquad (5.4)
\end{aligned}$$

This proves (5.2).

Now, for any $\mu > 0$, applying chain rule and taking the derivative of (4.1) with respect to μ we have

$$\begin{aligned} \phi'(\mu, x) &= \eta'(\mu, x(\mu)) - \eta'(\mu, x) + \nabla\eta(\mu, x(\mu))^T x'(\mu) \\ &= \eta'(\mu, x(\mu)) - \eta'(\mu, x). \end{aligned} \tag{5.5}$$

From (5.5) applying the mean-value theorem we obtain

$$\begin{aligned} |\phi'(\mu, x)| &= \left| \int_0^1 \{\nabla\eta(\mu, x - (1-\tau)\tilde{\Delta}x)^T\}'\tilde{\Delta}x \; d\tau \right| \\ &\leq \int_0^1 \left[-\tilde{\Delta}x^T \nabla^2\eta(\mu, x - (1-\tau)\tilde{\Delta}x)\tilde{\Delta}x \right]^{1/2} \\ &\quad \left[-\{\nabla\eta(\mu, x - (1-\tau)\tilde{\Delta}x)^T\}'[\nabla^2\eta(\mu, x - (1-\tau)\tilde{\Delta}x)]^{-1}\{\nabla\eta(\mu, x - (1-\tau)\tilde{\Delta}x)^T\}' \right]^{1/2} d\tau. \end{aligned} \tag{5.6}$$

From (5.6), and using (3.17) from the proof of Lemma 3.2, and Proposition 5.1(i) we get

$$\begin{aligned} |\phi'(\mu, x)| &\leq \int_0^1 \frac{\sqrt{\tilde{\Delta}x^T \nabla^2\eta(\mu, x)\tilde{\Delta}x}}{1 - \tilde{\delta} + \tau\tilde{\delta}} \sqrt{\frac{p + Kr}{\mu}} \; d\tau \\ &\leq \int_0^1 \frac{\sqrt{\mu}\tilde{\delta}}{1 - \tilde{\delta} + \tau\tilde{\delta}} \sqrt{\frac{p + Kr}{\mu}} \; d\tau = -\sqrt{p^2 + Kr^2} \ln(1 - \tilde{\delta}). \quad \square \end{aligned}$$

Lemma 5.4 *Let $\mu > 0$ and $x \in \mathcal{F}^0$ be such that $\tilde{\delta} < 1$, where $\tilde{\delta}$ is as defined in Lemma 5.3. Let $\mu^+ = \gamma\mu$ with $\gamma \in (0, 1)$. Then,*

$$\eta(\mu^+, x(\mu^+)) - \eta(\mu^+, x) \leq O(p + Kr)\mu^+.$$

Proof. Differentiating (5.5) we obtain

$$\phi''(\mu, x) = \eta''(\mu, x(\mu)) + \nabla\eta'(\mu, x(\mu))^T x'(\mu) - \eta''(\mu, x). \tag{5.7}$$

Now we will bound the two terms on the right hand side of (5.7) separately. From the definition of $\eta(\mu, x)$ in (2.5) it can be easily seen that for $\mu > 0$ and $x \in \mathcal{F}^0$, $\eta''(\mu, x) = \sum_{i=1}^K \rho_i''(\mu, x)$. Differentiating $\rho_i(\mu, x)$ and using (3.15) we obtain

$$\begin{aligned} \rho_i'(\mu, x) &= d_i^T y_i' + \ln\det S_i + \mu s_i^{-T} s_i' \\ &= \ln\det S_i + (-d_i + W_i^T\lambda_i)^T R_i^{-1} W_i^T s_i^{-1} \\ &= \ln\det S_i. \end{aligned} \tag{5.8}$$

Differentiating (5.8) once more and using (3.15) for $\mu > 0$ and $x \in \mathcal{F}^0$ we get

$$\begin{aligned} \rho_i''(\mu, x) &= s_i^{-T} s_i' \\ &= s_i^{-T} W_i R_i^{-1} W_i^T s_i^{-1} \\ &= s_i^{-T} Q_i^{-1} P_i Q_i^{-1} s_i^{-1} \\ &\leq s_i^{-T} Q_i^{-2} s_i^{-1} \\ &= s_i^{-T} \lambda_i^{-1} = \frac{r}{\mu}, \end{aligned}$$

and thus

$$\eta''(\mu, x(\mu)) \leq \frac{Kr}{\mu}. \tag{5.9}$$

Differentiating the optimality condition of the first stage problem (2.5) we observe

$$x(\mu)' = -[\nabla^2 \eta(\mu, x(\mu))]^{-1} \nabla \eta'(\mu, x(\mu)). \tag{5.10}$$

Hence, we have

$$\begin{aligned} \nabla \eta'(\mu, x(\mu))^T x'(\mu) &= -\nabla \eta'(\mu, x(\mu))^T [\nabla^2 \eta(\mu, x(\mu))]^{-1} \nabla \eta'(\mu, x(\mu)) \\ &\leq \mu^{-1}(p + Kr). \end{aligned} \tag{5.11}$$

In the last inequality we used (3.17) which is valid for any $\mu > 0$ and $x \in \mathcal{F}^0$. Combining (5.9) and (5.11) we have

$$\phi''(\mu, x) \leq \mu^{-1}(p + 2Kr). \tag{5.12}$$

Now in view of Lemma 5.3 and (5.12) we have

$$\begin{aligned} \phi(\mu^+, x) &= \phi(\mu, x) + \phi'(\mu, x)(\mu^+ - \mu) + \int_\mu^{\mu^+} \int_\mu^\tau \phi''(\nu)\ d\nu\ d\tau \\ &\leq \mu \left[\frac{\tilde{\delta}}{1 - \tilde{\delta}} + \ln(1 - \tilde{\delta}) \right] - \sqrt{p + Kr} \ln(1 - \tilde{\delta})(\mu - \mu^+) \\ &\quad + (p + 2Kr) \int_\mu^{\mu^+} \int_\mu^\tau \nu^{-1}\ d\nu\ d\tau \\ &\leq \mu \left[\frac{\tilde{\delta}}{1 - \tilde{\delta}} + \ln(1 - \tilde{\delta}) \right] - \sqrt{p + Kr} \ln(1 - \tilde{\delta})(\mu - \mu^+) \\ &\quad + (p + 2Kr) \ln \gamma^{-1}\ (\mu - \mu^+). \end{aligned} \tag{5.13}$$

Since γ and $\tilde{\delta}$ are absolute constants (5.13), and the fact that $\eta(\mu, x)$ is a strictly convex function in μ (implying $\eta''(\mu, x) > 0$) we have a proof of this lemma. □

Note that Lemma 5.3 and Lemma 5.4 require $\tilde{\delta}$ be less than one. However, we cannot evaluate $\tilde{\delta}$ since we do not explicitly know the points $x(\mu)$ forming the central path. Nonetheless we can evaluate δ and $\tilde{\delta}$ is proportional to δ, as shown in the following Lemma.

Lemma 5.5 *For any given* $\mu > 0$, $x \in \mathcal{F}^0$ *let* Δx *be the Newton direction defined in (4.2) and* $\tilde{\Delta} x := x - x(\mu)$. *We denote*

$$\delta := \delta(\mu, x) = \sqrt{-\frac{1}{\mu}\Delta x^T \Omega \Delta x} \quad \textit{and} \quad \tilde{\delta} := \tilde{\delta}(\mu, x) = \sqrt{-\frac{1}{\mu}\tilde{\Delta} x^T \Omega \tilde{\Delta} x},$$

where $\Omega := \nabla^2 \eta(\mu, x)$. *If* $\delta \leq 1/6$ *then*

$$\frac{2}{3}\delta \leq \tilde{\delta} \leq 2\delta. \tag{5.14}$$

Proof. Let $H := \nabla^2 \eta(\mu, x)$, $g := \nabla \eta(\mu, x)$. We denote $\overline{g} := g + H\tilde{\Delta} x$. Note that

$$\tilde{\Delta} x = -\Delta x + H^{-1}\overline{g}. \tag{5.15}$$

By applying the triangle inequality to (5.15) we obtain

$$\tilde{\delta} \leq \delta + \sqrt{-\frac{1}{\mu}\overline{g}^T H^{-1} \overline{g}}. \tag{5.16}$$

It is straightforward to verify that

$$-\overline{g} H^{-1} \overline{g} = \max\{h^T H h - 2h^T \overline{g} \mid h \in \mathbb{R}^n\}.$$

Now in consideration of Proposition 5.1 (ii) we have

$$\begin{aligned}
-h^T \overline{g} &= \int_0^1 h^T [\nabla^2 \eta(\mu, x) - \nabla^2 \eta(\mu, x - (1-\tau)\tilde{\Delta} x)]\tilde{\Delta} x d\tau \\
&\leq \int_0^1 [(1 - (1-\tau)\tilde{\delta})^{-2} - 1] d\tau \ \sqrt{\tilde{\Delta} x^T \nabla^2 \eta(\mu, x) \tilde{\Delta} x}\sqrt{h^T \nabla^2 \eta(\mu, x) h} \\
&= \frac{\sqrt{\mu}\tilde{\delta}^2}{1 - \tilde{\delta}}\sqrt{-h^T H h}.
\end{aligned}$$

Hence,

$$\begin{aligned}
-\overline{g} H^{-1} \overline{g} &\leq max\{h^T H h + \frac{2\sqrt{\mu}\tilde{\delta}^2}{1-\tilde{\delta}}\sqrt{-h^T H h} \mid h \in \mathbb{R}^n\} \\
&= \frac{\mu \tilde{\delta}^4}{(1-\tilde{\delta})^2}.
\end{aligned} \tag{5.17}$$

Combining (5.16) and (5.17) we obtain

$$\tilde{\delta} \leq \delta + \frac{\tilde{\delta}^2}{(1-\tilde{\delta})}. \tag{5.18}$$

When $\delta \leq \frac{1}{6}$, the quadratic inequality (5.18) implies $\tilde{\delta} \leq 2\delta$. The condition $\delta \leq 1/6$ is eventually reached since the inner iterations of the long-step algorithm converge.

From (5.16), exchanging positions of Δx and $\tilde{\Delta} x$ and following the above steps we can derive

$$\delta \leq \tilde{\delta} + \frac{\tilde{\delta}^2}{(1-\tilde{\delta})},$$

which in turn implies

$$\delta \leq \frac{3}{2}\tilde{\delta},$$

since $\tilde{\delta} \leq 2\delta \leq \frac{1}{3}$. $\square$

Lemma 5.1 implies that each inner iteration decreases the value of η by at least $\mu[\delta - \ln(1+\delta)]$. Therefore, in view of Lemma 5.1 and Lemma 5.4 it is clear that after reducing μ by a factor $\gamma \in (0,1)$, at most $O(p + Kr)$ Newton iterations will be needed for recentering. In the long-step version of our algorithm we need to update the barrier parameter μ no more than $O(\ln \mu^0/\epsilon)$ times.

Theorem 4.2 follows from Lemma 5.1 (ii), Lemma 5.4 and Lemma 5.5.

6. Concluding Remarks

We have described and analyzed a decomposition algorithm that follows the primal central trajectory in the first stage. At each iteration using optimal dual solutions of the second stage barrier problems, it generates gradient and Hessian information for the first stage problem and takes a Newton step in the primal space. Although the algorithm is attractive from the decomposition point of view, it has several limitations that need further work to develop practical implementations. These include: (i) generating a good starting point; (ii) development of practical step length selection procedure in the primal space; (iii) a practical strategy for reducing μ in our context; (iv) adaptive addition of scenarios; (v) computation of approximate solutions of the second stage problems, and (vi) a proper choice of ϵ to terminate the algorithm. These issues are addressed in a companion paper [8].

Bibliography

[1] O. Bahn, O. du Merle, J.-L. Goffin and J.P. Vial (1995), "A cutting plane method from analytic centers for stochastic programming," *Mathematical Programming* 69, 45-73.

[2] J. R. Birge, C. J. Donohue, D. F. Holmes, and O. G. Svintsiski (1996), "A parallel implementation of the nested decomposition algorithm for multistage stochastic linear programs," *Mathematical Programming* 75, 327- 352.

[3] J. R. Birge and D. F. Holmes (1992), "Computing block-angular Karmarkar projections with applicications to stochastic programming," *Computational Optimization and Applications* 1, (1992) 245-276.

[4] T. Carpenter, I. Lustig and J. Mulvey (1991), "Formulating stochastic programs for interior point methods," *Operations Research* 39, 757-770.

[5] J. Czyzyk, R. Fourer and S. Mehrotra (1995), "A study of the augmented system and column splitting approaches for solving two-stage stochastic linear programs by interior point methods," *ORSA Journal on Computing* 7, 474-490.

[6] J. Linderoth, A. Shapiro, S. Wright (2002), "The empirical behavior of sampling methods for stochastic programming", Published electronically in: www.optimization-online.org.

[7] S. Mehrotra and M.G. Özevin (2004), "Decomposition-Based Interior Point Methods for Two-Stage Stochastic Convex Quadratic Programs with Recourse," published on line http://www.speps.info/

[8] S. Mehrotra and M.G. Özevin (2004), "On the implementation of interior decomposition algorithms for stochastic semi-definite programs," (in preparation).

[9] J. E. Nesterov and A. S. Nemirovsky (1994), *Interior-Point Polynomial Algorithms in Convex Programming*, Studies in Applied Mathematics, SIAM.

[10] M. Ramana, L. Tunçel, and H. Wolkowicz (1997), "Strong duality for semidefinite programming," *SIAM Journal on Optimization*, 7(3), 641-662.

[11] J Renegar (2001), *A Mathematical view of Interior-Point Methods in Convex Programming*, MPS-SIAM Series on Optimization, SIAM.

[12] R. T. Rockafellar and R. J-B. Wets (1986), "A Lagrangian finite generation technique for solving linear-quadratic problems in stochastic programming," *Mathematical Programming Study* 28, 63-93.

[13] R. T. Rockafellar and R. J-B. Wets (1991), "Scenarios and policy aggregation in optimization under uncertainty," *Math. of Oper. Res.* 16, 119-147.

[14] R. M. Van Slyke and R. J. Wets (1969), "L-Shaped linear programs with applications to optimal control and stochastic linear programming," *SIAM Journal of Applied Mathematics* 17, 638-663.

[15] H. Wolkowicz, R. Saigal, and L. Vandenberghe (2000) (ed.), *Handbook of Semidefinite Programming: Theory, Algorithms,and Applications*, Kluwer.

[16] G. Zhao (2001), "A log-barrier method with Benders decomposition for solving two-stage stochastic linear programs," *Mathematical Programming* Ser. A 90, 507-536.